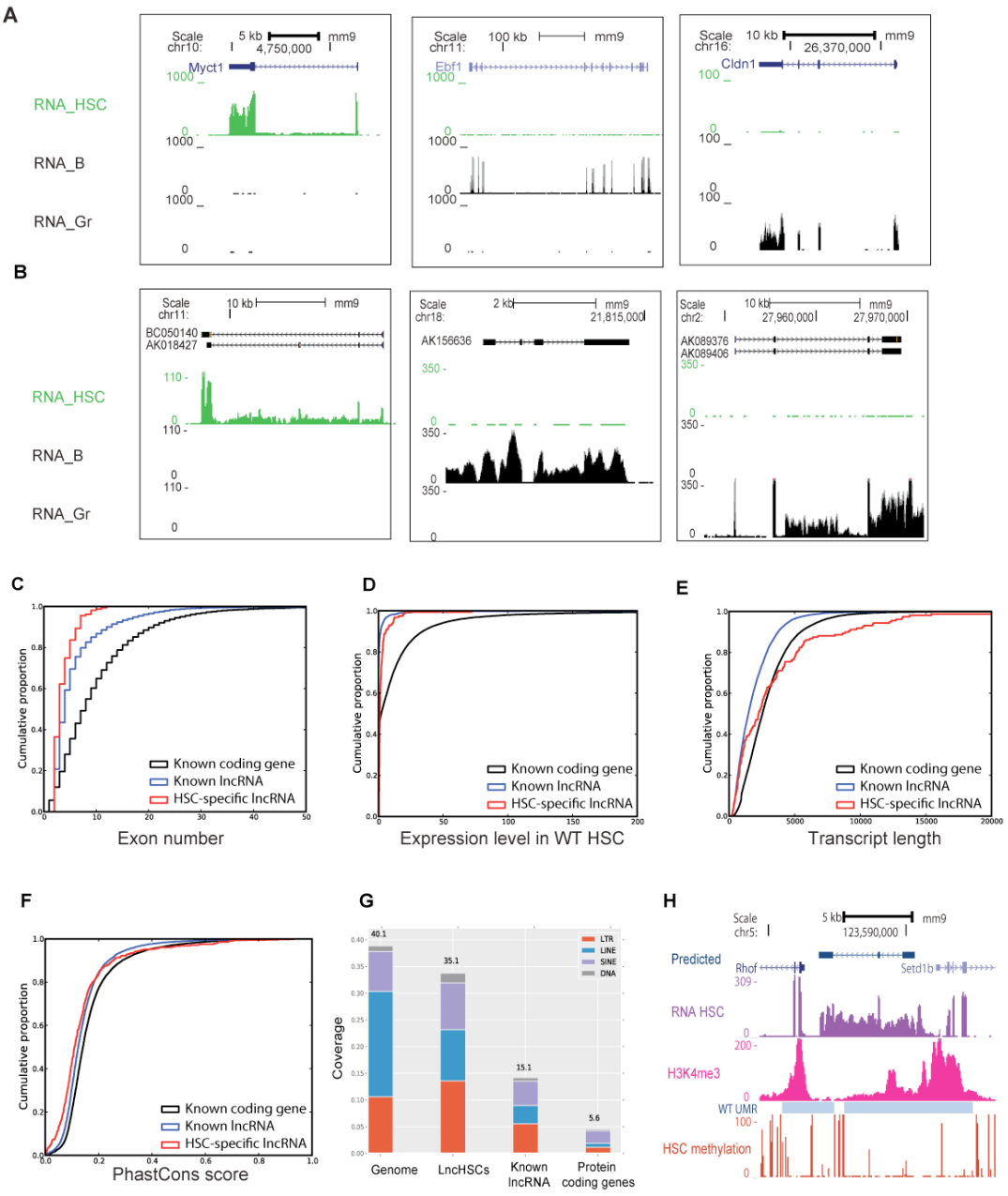

Figure S1. Similarity of HSC-specific lncRNAs to protein coding genes. (A) UCSC genome browser depicting HSC, B cell and Granulocytes fingerprint genes *Myct1*, *Ebf1* and *Cdln1* expression by RNA-seq. (B) UCSC genome browser depicting HSC–, B-cell– and Granulocyte–specific annotated lncRNAs AK018427, AK156636 and AK089406 by RNA-seq. Features of 159 HSC-specific lncRNAs (LncHSCs, red) compared to 20,156 known coding genes (black) and 2,614 known lncRNAs (blue) for exon number (C), expression levels (D), transcript length (E) and Phastcon score (F). (G) Coverage of different transposable elements in genome, LncHSCs, known lncRNAs and protein coding genes. Here TE includes LTR, LINE, SINE, and DNA repeats. (H) UCSC browser track showing expression (purple), H3K4me3 signal (pink) and DNA methylation (red) for one LncHSC. UMR is marked by light blue bar. Related to Figure 1.

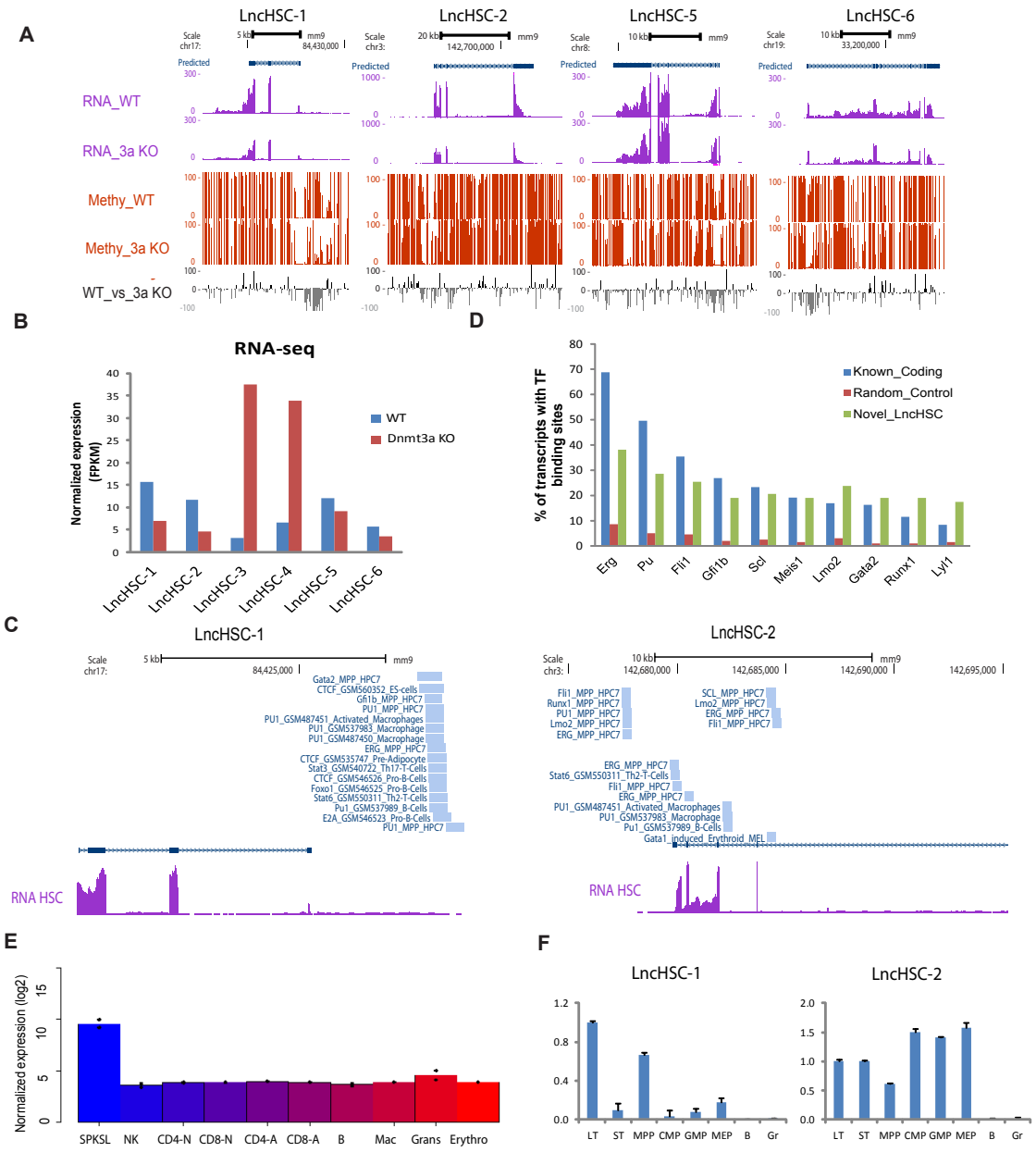

Figure S2. Validation of selected LncHSCs. (A) UCSC browser track showing expression (purple) and DNA methylation (red) for LncHSC-1, 2 , 5 and 6 in WT and *Dnmt3a* KO (3a KO) HSCs. Grey bars show differential methylation between WT and 3a KO HSCs. (B) Expression levels of LncHSC1-6 in WT and *Dnmt3a* KO HSC by RNA-seq. The y- axis indicates the value of FPKM. (C) UCSC browser track showing transcription factor binding sites (light blue bars) on the promoter of LncHSC-1 and lncHSC-2. (D) Percentage of protein-coding genes or LncHSCs with specific TF binding sites at their TSS±5kb region. (E) LncHSC-2 expression in HSC (SPKSL), natural killer cells (NK), naïve CD4 and CD8 T cells (CD4-N and CD8-N), activated CD4 and CD8 T cells (CD4-A and CD8A), B cell, Macrophage (Mac), Granulocytes (Grans) and Erythrocytes (Erythro) by microarray. The Y-axis represents expression level (log2 scale). (F) LncHSC-1/2 expression in LT-HSC, ST-HSC, MPP, CMP, GMP, MEP, B and Gr cells. Error bars represent Mean ± SEM. (n=3). Related to Figure 2.

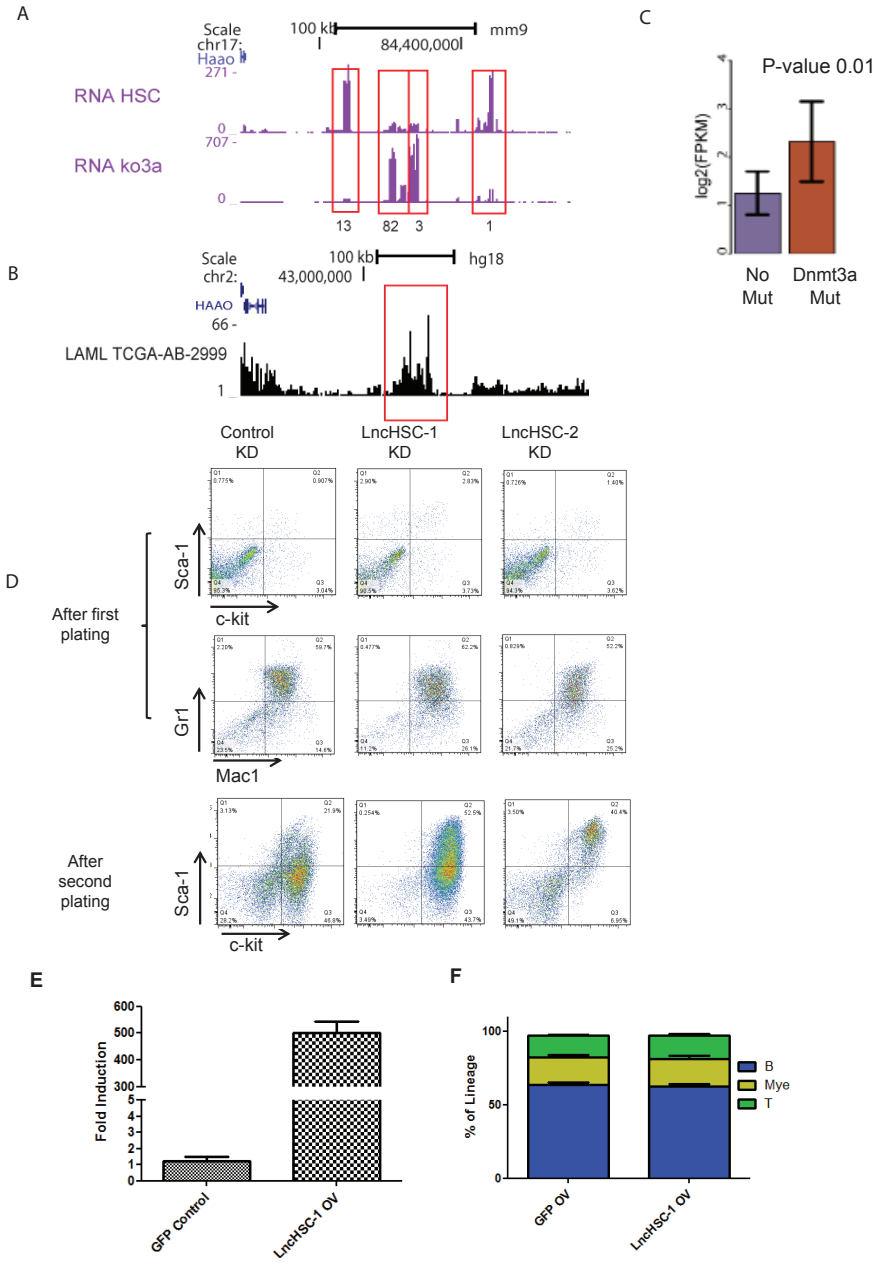

Figure S3. LncHSCs control HSC *in vitro* differentiation. (A) UCSC browser track showing 4 LncHSCs (LncHSC-1, 3, 82 and 13) expressed between *Haao* and *Zfp36l2* genes in mouse WT and *Dnmt3a* KO HSCs. (B) UCSC browser track showing several novel transcripts expressed between HAAO and ZFP36L2 genes in human TCGA patients.The most abundantly expressed transcript is marked by red box. (C) Bar graph showing expression level of the most abundantly expressed novel transcript between HAAO and ZFP36L2 genes in 122 LAML samples, among which 27 with DNMT3A mutation. (D) Flow cytometry analysis of the cells in the colonies with LncHSC knockdown after first plating and second plating. (E) RT-PCR to confirm that LncHSC-1 over-expression in vivo. Error bars represent Mean ± SEM. (n=3). (F) PB analysis post-transplant 16 weeks. The percentage of cells of the indicated lineages within the CD45.2^{+}GFP^{+} cell compartment is shown. Myeloid cells (Mye) were defined as Gr1^{+} and Mac1^{+}, B-cells (B) are B220^{+}, T-cells (T) are CD4^{+} and CD8^{+}. Error bars: Mean ± SEM. (n=10). Related to Figure 3.

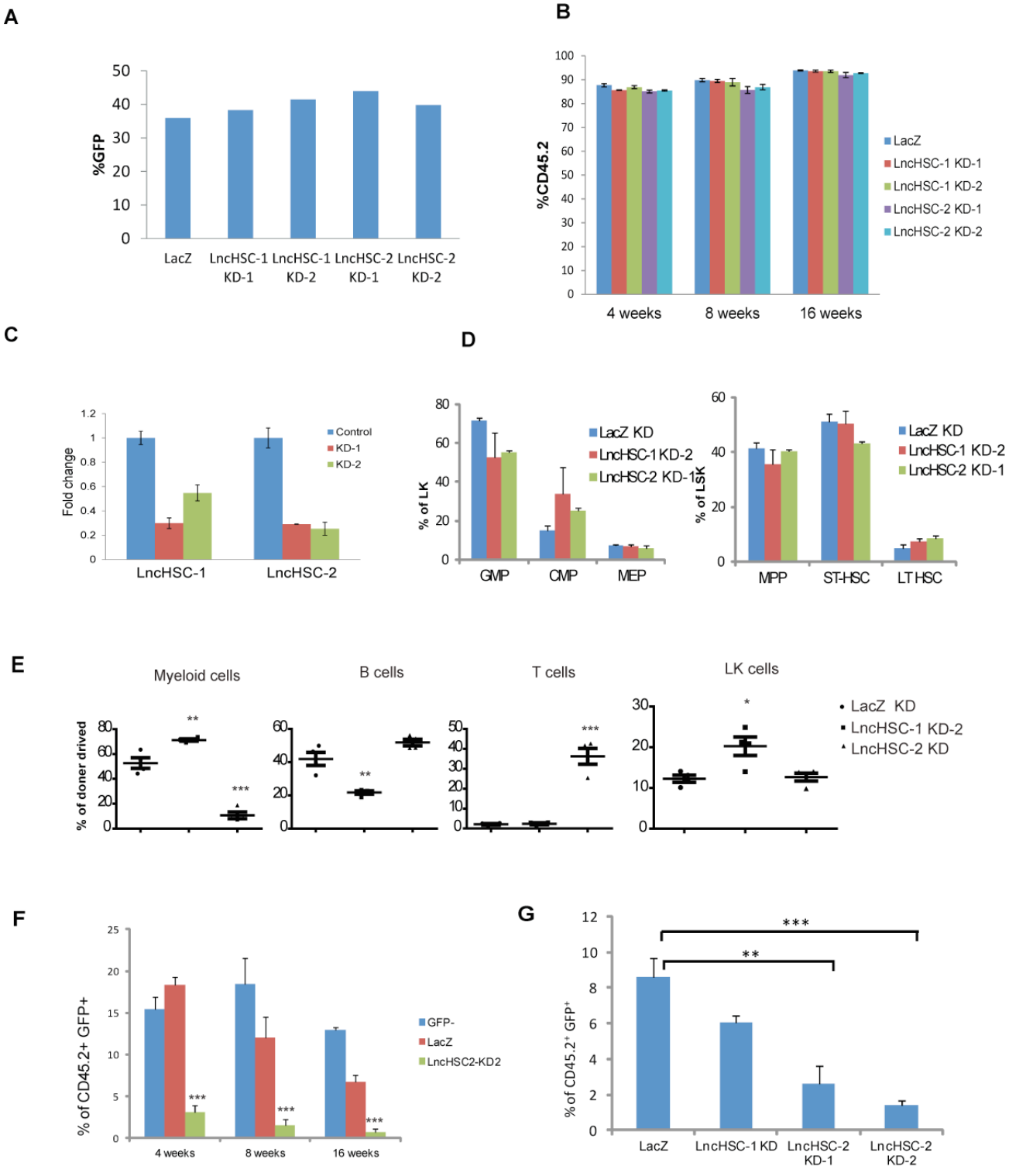

Figure S4. Luo *et al*

Figure S4. LncHSCs regulate HSC *in vivo* differentiation. (A) Transduction efficiency of miRNA retrovirus in Sca-1^{+} cells determined by percent of GFP+ population after in vitro culture for two days. (B) Engraftment of donor cells (CD45.2^{+}) in the recipient mouse peripheral blood after transplantation 16 weeks. (n=10 for Control KD, n=5-8 for LncHSCs KD). (C) RT-PCR results showing that LncHSC-1 and LncHSC-2 are knocked-down in the CD45.2GFP+-KSL population after transplantation 20 weeks. Error bars represent Mean ± SEM. (n=3). (D) Progenitor analysis of the CD45.2GFP+ cells in the bone marrow after transplantation 20 weeks. Mean ± SEM. (n=3). (E) Bone marrow FACS analysis showing frequencies of GFP+ progenitor and differentiated cell populations in mice 20 weeks after primary transplantation. Mean ± SEM. (n=3). *P < 0.05, **P < 0.01, ***P < 0.001. (F) Peripheral blood analysis of donor-derived transduced (GFP+) cells after secondary transplantation with LncHSC-2-KD2. *** P <0.001. Error bars represent Mean ± SEM. (n=4-5). (G) Analysis of bone-marrow of secondary transplant recipients for donor-derived cells. Contribution of retrovirally-transduced donor cells (CD45.2^{+}GFP^{+}) to recipient mouse bone marrow after secondary-transplantation 16 weeks. *** P <0.001, ** P< 0.01 Error bars represent Mean ± SEM (n=3-6). Related to Figure 4.

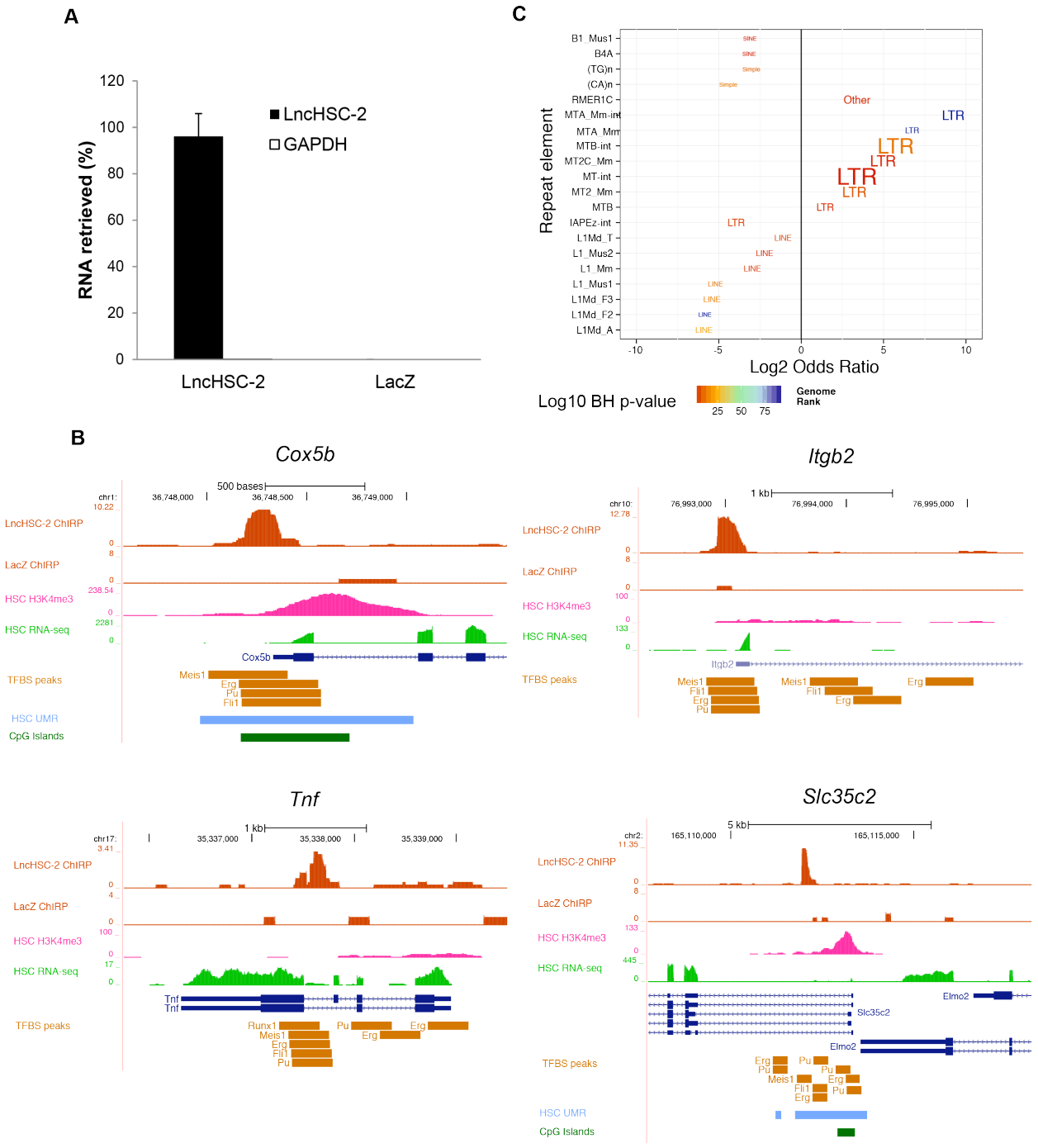

Figure S5. LncHSC-2 ChIRP-seq of HPC5 hematopoietic progenitor cells. (A) RT-PCR showed that ChIRP retrieves 95% of cellular LncHSC-2 RNA and undetectable GAPDH RNA. LacZ probes are used as negative controls and retrieve neither RNAs. Error bars represent Mean ± SEM (n=3). (B) LncHSC-2 occupancy at the genes *Cox5b*, *Itgb2*, *Tnf*, and *Slc35c2*. LncHSC-2 and LacZ control ChIRP-seq signal density tracks generated by MACS2 representing the fragment pileup signal per million reads. Additional overlaid tracks are HSC H3K4me3, RNA-seq, undermethylated regions (Jeong et al., 2014) and hematopoietic lineage TF binding sites (Wilson et al., 2010). (C) LncHSC-2 binding sites are enriched for enriched for LTR family of repeats. After performing peak calling with unique and secondary alignments,

we assessed the enrichment or depletion for LncHSC-2 binding compared to LacZ negative control across repetitive sequence elements (taken from RepeatMasker mm9 mouse assembly) by two-tailed Fisher's exact test with multiple testing correction. The names of significant repeats are given on the y-axis and labeled on the plot by their Class (LINE, LTR, Other, Simple, or SINE). Label sizes are proportionate to the descending rank of the repeat's frequency in the genome (smaller label = higher rank = more frequent, larger label = lower rank = less frequent). Label color gradient represents the –log10 Benjamini-Hochberg corrected p-value. The x-axis values are the log2 Odds Ratio reported by Fisher's exact test. Repeat classes and frequencies in ChIRP-seq peaks in Table S6. Related to Figure 5.

Supplemental Tables

Please see accompanying Excel files

Table S1. Hematopoietic lncRNAs identified by RNA-seq; Related to Figure 1.

Table S2. Expression data after knockdown of specific LncHSCs; Related to Figure 3.

Table S3. LncHSC-2 peaks identified by ChIRP-seq; Related to Figure 5.

Table S4. ChIRP-seq motif analysis; Related to Figure 5B

Table S5. Factors and sequence features enriched in LncHSC-2 binding sites; Related to Figure 5C

Table S6. Significantly enriched or depleted repeat elements in LncHSC-2 binding sites; Related to Figure 5

EXTENDED EXPERIMENTAL PROCEDURES

Hematopoietic Stem Cell Purification and Flow Cytometry

All animal procedures were IUCAC approved and conducted in accordance with institutional guidelines. Mice were housed in a specific-pathogen-free facility and fed autoclaved acidified water and mouse chow ad libitum. Whole bone marrow cells were isolated from femurs, tibias, pelvis and humerus. For antibody staining, cells were suspended at a concentration of 10^8 cells/ml and incubated in 4°C for 15 minutes with the desired antibodies. B cells ($B220^+$) and Gr cells ($Gr1^+$) were sorted from 12-month-old mice after antibody staining. For HSCs, SP staining was performed with Hoechst 33342 (Sigma) as previously described (Goodell et al., 1996). Briefly, whole bone marrow cells were resuspended in staining media at 10^6 cells/mL and incubated with 5 μg/ml Hoechst 33342 for 90 minutes at 37°C. Magnetic enrichment was performed with anti-mouse c-Kit (CD117) microbeads (Miltenyi Biotec, Germany) on an AutoMACS (Miltenyi Biotec, Germany). Post-enrichment, the positive cell fraction was labeled with antibodies to identify HSCs (SP^+ Lineage (CD3, CD4, CD8, B220, Gr1, Mac1 and T119)$^-$ Sca-1^+ c-Kit^+ $CD150^+$). All antibodies were obtained from BD Biosciences (San Jose, CA) or eBioscience (San Diego, CA) and used at 1:100 dilutions. Cell sorting was performed on a MoFlo cell sorter (Dako North America, Carpinteria, CA) or Aria II (BD Biosciences, San Jose, CA) and analysis performed on a LSRII (BD Biosciences, San Jose, CA). This population of HSCs is also $CD34^-Flk2^-CD48^-$ (Mayle et al., 2013), and single cells from this pool are capable of reconstitution of all blood cell types (Challen et al., 2010).

RNA-sequencing

~70,000 HSCs, 1 million B cells and Granulocytes were FACS sorted. RNA was isolated with the RNeasy Micro kit (Qiagen), including the DNase I (Qiagen) on-column digestion. Paired end libraries were generated by using Illumina TruSeq RNA sample preparation kit. Illumina HiSeq was used for sequencing with a paired-end sequencing length of 100bp.

RNA-Seq data alignment

Paired-end 100 bp reads were sequenced. The last 20 bases were trimmed due to average low quality. The alignment was performed by RUM (Grant et al., 2011), which first mapped reads to the genome and transcriptome by Bowtie, and then used Blat to re-map those initially unmapped reads to the genome. The information from the two rounds of mappings was merged. The multiply-mapped reads were discarded. The gene annotations used for transcriptome alignment include RefSeq, UCSC known Gene and Ensemble gene models. The gene expression, FPKM value, was calculated by counting the reads matching the exons of each gene.

Transcriptome assembly

There are two biological replicates for each of the HSC samples, including 4 month HSC, 12 month HSC, 24 month HSC and *Dnmt3a* KO HSC. For each replicate, the RUM aligned reads were sent to Cufflinks for transcriptome assembly with default parameters. The strand of spliced reads was estimated from splicing motifs. Intron reads are removed before assembly. In order to obtain a highly confident transcriptome, we required splicing evidence from the same location on both replicates. The transcriptome from each sample was hence regarded as an accurate

reflection of the HSC transcriptome at various conditions. All 4 such HSC transcriptomes were then merged, and a final comprehensive HSC transcriptome was obtained.

Determination of coding probability

We determined novel genes as lncRNA based on coding potential calculated from CPAT software (Coding Potential Assessment Tool) (Wang et al., 2013). As the coding potential cutoff is trained using known coding and non-coding RNAs of mouse to be 0.44, CPAT shows 97% sensitivity and specificity for mouse lncRNA estimation.

Annotation of novel genes by epigenetic and genetic data

The novel genes are annotated by histone modifications H3K4me3 and H3K36me3, by low DNA methylation regions, and by 10 important HSC transcription factors' binding sites. Here the histone modification data for wild type HSC is from GSE47765 and DNA methylation data for wild type HSC is from GSE47815. The peak regions for 10 TFs are from http://hscl.cimr.cam.ac.uk/ChIP-Seq_Compendium/ChIP-Seq_Compendium2.html. Repeat masker data and mouse ESTs are obtained from the UCSC genome browser and are annotated for overlap with the novel genes.

Calculation of tissue specificity

The specificity score for each gene is calculated by the rank difference in the two FPKM distributions. The gene is specific, if its rank is higher in one distribution by more than 20%.

Data accession and UCSC Genome Browser Track hub

The WT 4-month and 24-month HSC RNA-Seq data are obtained from GSE47817. The 12-month HSC and *Dnmt3a* knockout HSC RNA-Seq data are obtained from (Jeong et al., 2014). The B-Cell RNA-Seq data is obtained from GSE50775. The Granulocyte RNA-Seq data and the knockout down of LncRNAs RNA-Seq data are deposited in NCBI GEO under accession number GSE53928. All the HSC RNA-Seq data and novel gene structures can be visualized on UCSC Genome Browser through the trackhub URL http://dldcc-web.brc.bcm.edu/lilab/ncRNA/tracks.txt. To visualize the data, go to the UCSC Genome Browser page for the mouse genome mm9, select "Track hub" under the "myData" tab, insert the URL, and click on the "Add Hub" button. One may also simply visit http://genome.ucsc.edu/cgi-bin/hgTracks?db=mm9&hubUrl=http://dldcc-web.brc.bcm.edu/lilab/deqiangs/ncRNA/tracks.txt to access this track hub.

ChIP-sequencing (ChIP-seq)

Chromatin Immunoprecipitation (ChIP) was performed as described previously (Dahl and Collas, 2008). Briefly, 50,000~100,000 HSCs, 1000,000 B cell and Gr cells were sorted and crosslinked with 1% formaldehyde at room temperature (RT) for 10 min, and the reaction was stopped by 0.125M glycine at RT for 5 min. Then the cells were washed once with ice cold PBS containing protease inhibitor cocktail (PIC; Roche) and the cell pellet was stored at −80°C. Cross-linked cells were thawed on ice and lysed in 50 µl Lysis buffer (10 mM Tris pH 7.5, 1mM EDTA, 1% SDS), then diluted with 150 µl of PBS/PIC, and sonicated to 200-500 bp fragments (Bioruptor, Diagenode). The sonicated chromatin was centrifuged at 4°C for 5 min at 13,000rpm to remove precipitated SDS. 180 µl was then transferred to a new 0.5 ml collection tube, and 180 µl of 2X RIPA buffer (20 mM Tris pH 7.5, 2 mM EDTA, 2%Triton X-100, 0.2% SDS, 0.2% sodium

deoxycholate, 200 mM NaCl/PIC) was added to recovered supernatants. A 1 /10 volume (36 µl) was removed for input control. ChIP-qualified antibodies (0.1 µg H3K4me3 Millipore 07-473, 0.3 µg H3K36me3 Abcam ab9050, 0.3 µg H3K4me1 Abcam ab8895 and 0.3 µg H3K27ac Abcam ab4729) were added to the sonicated chromatin and incubated at 4°C overnight. Following this, 10 µl of protein A magnetic beads (Dynal, Invitrogen) previously washed in RIPA buffer were added and incubated for an additional 2 hours at 4°C. The bead: protein complexes were washed three times with RIPA buffer and twice with TE (10 mM Tris pH 8.0/1 mM EDTA) buffer. Following transfer into new 1.5 ml collection tube, genomic DNA was eluted for 2 hours at 68 °C in 100 µl Complete Elution Buffer (20 mM Tris pH 7.5, 5 mM EDTA, 50 mM NaCl, 1% SDS, 50 µg/ml proteinase K), and combined with a second elution of 100 µl Elution Buffer (20 mM Tris pH 7.5, 5 mM EDTA, 50 mM NaCl) for 10 min at 68 °C. ChIPed DNA was purified by MinElute Purification Kit (Qiagen) and eluted in 12 µl elution buffer. ChIPed DNA were successfully made library using ThruPLEX-FD preparation kit without extra amplification (Rubicon, Ann Arbor, MI). Sequencing was performed according to the manufacturer's protocol on a HiSeq 2000 (Illumina).

For ChIP-qPCR of E2A binding, about 20,000-50,000 Sca-1+ cells transduced by control or LncHSC-2 KD construct were used. 0.5 µg E2A antibody (Santa Cruz Biotechnology; SC-349) was used for each IP. The primers used for the three LncHSC-2 binding sites were shown in Table 1. Negative control is a region not containing any E2A or LncHSC-2 binding sites.

Gene	Forward	Reverse
Nln	GGTATCGCGTCTGATCCTGT	AAACGGTGCCACACATCTTT
Slc35c2	CCCCACTCCCTTACCTCAGT	CGCTAGGTCTCTGAGGAGGA
Itgb2	GTGTAAAGGGGCTGAGGTGA	CTGGCTGTGTGGTTGTGTCT
Control	GTGAAGGCGAGACGAAAAAG	CTGCACTCAAGGGACTCCTC
Table 1: Primers for ChIP-qPCR.		

Peak calling of histone modification data

The ChIP-seq reads are mapped to mouse genome mm9 using SOAP2 (Li et al., 2009) by allowing at most 2 mismatches for 50bp long short reads and at most 4 mismatches for 100bp long short reads. Only uniquely mapped reads were retained. To remove duplicate reads resulting from the PCR amplification, at most 2 duplicate reads were allowed for each biological replicate. The number 2 is based on Poisson P-value cutoff of $1x10^{-5}$ determined by the total number of reads with respect to the theoretical mean coverage across the genome. The uniquely mapped and duplicate removed reads from each biological replicate are fed as a treatment file into the MACS program (Zhang et al., 2008), to find the enriched regions, or "peaks". Peaks are regions with enrichment of treatment reads compared to control reads, which are just sonicated DNA fragments without ChIP pull down. The p-value cutoff for MACS is E-8. Peaks from all biological replicates of a specific sample are merged to form the final set of peaks for this specific sample.

Retrovirus over-expression

LncHSC-1 (AK039852) was subcloned into MSCV-RFB-IRES-GFP vector using Gateway recombination. Briefly, cDNA for AK039852 was PCR amplified and inserted into the pENTR-D/TOPO vector (Invitrogen) by TOPO cloning. Correct clones were sequence verified and recombined into an MSCV-RFB-IRES-GFP vector (containing attR recombination sites) using LR clonase enzyme mixture (Invitrogen) to produce retroviral vectors. MSCV-IRES-GFP was used as a control vector in all experiments. Viruses were packaged by cotransfection with pCL-Eco into 293T cells. Viral supernatants were collected 48-hours post-transfection and viral titers determined using 3T3 cells.

For retroviral transduction of hematopoietic progenitors, donor mice were treated with 5-fluorouracil (150mg/kg; American Pharmaceutical Partners, Schaumburg, IL) six days prior to bone marrow harvest. Whole bone marrow was enriched for Sca-1^+ cells using magnetic enrichment (AutoMACS; Miltenyi Biotec) and adjusted to a concentration of 5 x 10^5 cells/ml in transduction medium, containing Stempro 34 (Gibco, Carlsbad, CA), nutrient supplement, penicillin/streptomycin, L-glutamine (2mM), mSCF (10ng/ml; R&D Systems Minneapolis, MN), mTPO (100ng/ml; R&D Systems), and polybrene (4mg/ml; Sigma). The suspension was spin-infected at 250 x g at room temperature for 2 hours, and cells were incubated for a further 1 hour at 37^{o}C before injection.

miRNA cloning and retrovirus transduction

We used software Block-iT RNAi Designer (Invitrogen) to design miRNAs targeting HSC novel transcript. The stem-loop hairpin produces a miRNA that 100% matches to the gene of interest and cleaves the target mRNA. Oligos targeting each novel transcript were successfully cloned by BLOCK-iT PolII miR RNAi Expression Vector Kit (Invitrogen). Oligos targeting lacz were provided by the kit and used as control in all experiments. Briefly, the synthetic double-stranded oligos were cloned into the vector, pcDNA 6.2-GW/EmGFP-miR. The stem-loop hairpin with GFP tag was then incorporated into the pDonor vector using BP clonase enzyme mixture (Invitrogen). The oligos were further recombined into the retroviral MSCV-RFB vector (containing attR recombination sites) using LR clonase enzyme mixture (Invitrogen). Viruses were packaged by cotransfection with pCL-Eco into 293T cells. Viral supernatants were collected 48-hours post-transfection and viral titers determined using 3T3 cells. All the oligos that successfully knockdown the targets are described below.

For retroviral transduction of hematopoietic progenitors, donor mice were treated with 5-fluorouracil (150mg/kg; American Pharmaceutical Partners, Schaumburg, IL) six days prior to bone marrow harvest. Whole bone marrow was enriched for Sca-1^+ cells using magnetic enrichment (AutoMACS; Miltenyi Biotec) and adjusted to a concentration of 5 x 10^5 cells/ml in transduction medium, containing Stempro 34 (Gibco, Carlsbad, CA), nutrient supplement, penicillin/streptomycin, L-glutamine (2mM), mSCF (10ng/ml; R&D Systems Minneapolis, MN), mTPO (100ng/ml; R&D Systems). The suspension was spin-infected at 250 x g at room temperature for 2 hours in the presence of polybrene (4 μg/ml). For in vivo transplantation, cells

were incubated for a further 1 hour at 37^{o}C. For *in vitro* assays, transduced cells were cultured in fresh transduction medium for a further two days.

Oligo sequences used were (target sequence underlined):

Control –F
5'-TGCTGAAATCGCTGATTTGTGTAGTCGTTTTGGCCACTGACTGACGACTACACATCAGCGATTT-3'

Control-R
5'- CCTGAAATCGCTGATGTGTAGTCGTCAGTCAGTGGCCAAAACGACTACACAAATCAGCGATTTC -3'

LncHSC-1-1F
5'- TGCTGTATTCAGGCAGAGATCACTAAGTTTTGGCCACTGACTGACTTAGTGATCTGCCTGAATA -3'

LncHSC-1-1R
5'- CCTG TATTCAGGCAGATCACTAAGTCAGTCAGTGGCCAAAACTTAGTGATCTCTGCCTGAATAC -3'

LncHSC-1-2F
5'- TGCTGATTGGATGAAACCTTGTGGGCGTTTTGGCCACTGACTGACGCCCACAATTTCATCCAAT -3'

LncHSC-1-2R
5'- CCTGATTGGATGAAATTGTGGGCGTCAGTCAGTGGCCAAAACGCCCACAAGGTTTCATCCAATC -3'

LncHSC-2-1F
5'- TGCTGATAGAAACTTGGCCGATGAAGGTTTTGGCCACTGACTGACCTTCATCGCAAGTTTCTAT -3'

LncHSC-2-1R
5'- CCTGATAGAAACTTGCGATGAAGGTCAGTCAGTGGCCAAAACCTTCATCGGCCAAGTTTCTATC -3'

LncHSC-2-2F
5'-TGCTGAATACATGCTGCCTCTGCTCAGTTTTGGCCACTGACTGACTGAGCAGACAGCATGTATT-3'

LncHSC-2-2R
5'-CCTGAATACATGCTGTCTGCTCAGTCAGTCAGTGGCCAAAACTGAGCAGAGGCAGCATGTATTC-3'

In Vitro Colony-Forming Assay
WT C57Bl/6 mice were intraperitoneally injected with 5-FU. After 6 days, Sca-1^{+} cells were isolated from bone marrow and cultured in StemPro®-34 SFM media (Life Technologies) supplemented with penicillin/streptomycin, 2mM L-Glutamine, TPO (100 ng/ml) and stem cell factor (SCF, 10 ng/ml). The suspension was spin-infected at 250 x g by miRNA retrovirus at room temperature for 2 hours in the presence of polybrene. After transduction for 48 hr, 200 KSL-GFP^{+} cells were sorted into 6-well plates containing Methocult 3434 medium (Stem Cell

Technologies, Vancouver, BC, Canada) supplemented with penicillin / streptomycin and cultured *in vitro* at 37°C. Colonies were scored 14 days after seeding. For second plating, colonies were recollected in PBS, and then recultured in Methocult 3434 at a density 2,000 cells/replicate. Colonies were scored and analyzed 14 days after seeding.

In vivo Transplantation

All mice were C57Bl/6 background, distinguished by CD45.1 or CD45.2 alleles. For bone marrow transplantation, recipient C57Bl/6 CD45.1 mice were transplanted by retro-orbital injection following a split dose of 10.5 Gy of lethal irradiation. 50,000 Sca-1$^+$ (CD45.2) donor cells were injected to the recipient mice. For secondary transplantation, 500 CD45.2+GFP$^+$ KSL cells from primary recipients were re-sorted 20 weeks after transplantation and mixed with 250,000 CD45.1 WBM cells, and injected into new lethally irradiated recipients.

Peripheral Blood Analysis

For peripheral blood analysis by flow cytometry, mice were bled retro-orbitally, the red blood cells were lysed, and each sample was incubated with the following antibodies on ice for 20 min: CD45.1-FITC, CD45.2-APC, CD4-Pacific Blue, CD8-Pacific Blue, B220-Pacific Blue, B220-PeCy7, Mac1-PeCy7, and Gr-1-PeCy7 as previously described (Mayle et al., 2013). Cells were then spun down and re-suspended in a propidium iodide solution, and analysis was accomplished on live cells with an LSRII (Becton Dickinson).

Bone Marrow Progenitor Analysis

The progenitor analysis was performed as previously described (Mayle et al., 2013). Whole bone marrow cells was resuspended in 2% FBS DMEM at 10^6 cells mL^{-1} and incubated with a final concentration of 5 µg mL^{-1} Hoechst 33342 for 90 min at 37°C. For antibody staining, cells were suspended in 2% FBS HBSS staining media and incubated in 4°C for 15 min with various combinations of the following antibodies (all 1:100 dilution); Pacific blue-conjugated Mac-1, Gr-1, CD4, CD8, B220 and Sca-1-PeCy7, c-Kit- AF750, CD16/32-PE, CD34-APC, Flk2-PE (eBioscience, San Diego).

Quantitative Real-Time PCR

RNA was isolated from FACS-sorted KSL-GFP$^+$ cells using the RNeasy Micro kit (Qiagen). First-strand cDNA was synthesized by SuperScript II reverse transcriptase (Invitrogen). The relative levels of target mRNAs were quantitated using SYBR Green and the ABI 7900HT real-time PCR system. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was quantitated by Taqman using primer from ABI (4310884E). Samples were normalized to GAPDH and fold-change was determined by the ΔΔCt method. The gene-specific primers for novel transcripts are as follows:

Gene	Forward Primer (5'-3')	Reverse Primer (5'-3')
LncHSC-1	ATGAAACCTTGTGGGCTTTG	AGGCATATCATCCTGCCATC
LncHSC-2	GGCCTTGGAGATGAGAGATG	GATGAAGCATCCCAAATTGAC

LncHSC-3	GGCCAGATGGAAGAAATTCA	AGAAGTCGCAGAGGGAAACA
LncHSC-4	ATGCCAGAGTTTCTGGGAGA	GAGATGAGGCCTCCATTGTG
LncHSC-5	AAATGCCTGGATCCTTTGTG	ATTCATGGCCTGCTCTTCAT
LncHSC-6	ATTCACAGCCTTTGGGACAC	ATTTCCGCGTGTAAATCCAG
Table 2. RT-PCR primers		

Fluorescence in situ hybridization (FISH)

Single molecule RNA fluorescence in situ hybridization (FISH) was performed using the QuantiGene ViewRNA ISH Cell Assay according to manufacturer's instruction (Affymetrix). Probes were designed based on lncRNA sequences determined by Cufflinks. Images were taken on API Deltavision Deconvolution Microscope (Applied precision).

ChIRP

ChIRP was performed as described previously (Chu et al., 2011). Briefly, for 600-bp long LncHSC-2 transcript, twelve 3'-Biotin labeled anti-sense oligos were designed and synthesized (6 odd and 6 even) by Eurofins Genomics. 6 probes for Lacz were selected from (Chu et al., 2011) as non-specific control (Table 2). 10 million of HPC5 cells were sorted and crosslinked with 1% glutaraldehyde at room temperature (RT) for 10 min, and the reaction was stopped by 0.125M glycine at RT for 5 min. Then the cells were washed once with ice cold PBS containing PMSF, protease inhibitor cocktail (PIC; Roche) and the cell pellet was stored at −80°C. Cross-linked cells were thawed on ice and lysed in 400 µl nuclear lysis buffer (10 mM Tris pH 7.5, 1mM EDTA, 1% SDS, add PMSF, PIC and Superase-in before use), and sonicated to 200-500 bp fragments (Bioruptor, Diagenode). The sonicated chromatin was centrifuged at 4°C for 5 min at 13,000rpm to remove precipitated SDS. 350 µl was then transferred to a new 1.5 ml collection tube, add two times volume (700 µl) of hybridization buffer (750mMNaCl, 1% SDS, 50 mM Tris 7.0, 1 mM EDTA, 15% Formamide, add PMSF, PIC and Superase-in fresh). Probes of 100 pmol were added to 1,050 µl of diluted chromatin, which was mixed by end-to-end rotation at 37^{o}C for 4 hr. Streptavidin magnetic C1 beads were washed three times in nuclear lysis buffer, blocked with 500 ng/ul yeast total RNA, and 1mg/ml BSA for 1 hr at room temperature, and washed three times again in nuclear lysis buffer before resuspended in its original volume. One hundred microliters washed/blocked C1 beads were added per 100 pmol of probes, and the whole reaction was mixed for another 30 min at 37 ^{o}C. Beads:biotin-probes:RNA:chromatin adducts were captured by magnets (Invitrogen) and washed five times with 1 ml wash buffer (2XSSC, 0.5% SDS, add PMSF fresh). At the last wash, separate 100 µl beads for RNA elution and 900 µl beads for DNA elution.

ChIRP RNA Elution

Beads were re-suspended in 100 µl RNA PK buffer (100 mM NaCl, 10 mM Tris 7.0, 1mM EDTA, 0.5% SDS) and 0.2 U/ µl Proteinase K (Invitrogen). pK treatment was carried out at 65^{o}C for 45 min, followed by boiling for 15 min, and trizol:chloroform extraction. Eluted RNA was subject to quantitative reverse-transcription PCR (QRTPCR) for the detection of enriched transcripts.

ChIRP DNA Elution

Beads were re-suspended in 100 µl DNA elution buffer (50 mM $NaHCO_3$, 1% SDS, 200 mM NaCl), and DNA was eluted with a cocktail of 100 ug/ml RNase A (Sigma-Aldrich) and 0.1U/ µl RNase H (Epicenter). RNase elution was carried out twice at 37°C with end-to-end rotation and eluent from both steps was combined. Eluted chromatin was pK treated with 0.2 U/ µl pK at 65°C for 45 min. ChIPed DNA was purified by MinElute Purification Kit (Qiagen) and eluted in 12 µl elution buffer. ChIPed DNA was successfully made library using ThruPLEX-FD preparation kit (Rubicon, Ann Arbor, MI). Sequencing was performed according to the manufacturer's protocol on a HiSeq 2500 (Illumina).

Gene	Odd	Even
LncHSC2_1	tttagggcttgtaggctact	
LncHSC2_2		tgtatttaagctgctgaggc
LncHSC2_3	gctggctcattgtggaaata	
LncHSC2_4		aagaggaaaccagacccaga
LncHSC2_5	gatgcaagtttccaaagtcc	
LncHSC2_6		acagaagaggctgcatacga
LncHSC2_7	cgatatccagtgtcagtatg	
LncHSC2_8		aatgagtcataccgcttctg
LncHSC2_9	acaatctgaaggagctcctt	
LncHSC2_10		gctagctggctttgaaagtg
LncHSC2_11	ctgctgatcagcttcatgaa	
LncHSC2_12		gcatcccaaattgacttgta
Lacz_1	ccagtgaatccgtaatcatg	
Lacz_2	agatgaaacgccgagttaac	
Lacz_3	ataatttcaccgccgaaagg	
Lacz_4	ttcatcagcaggatatcctg	
Lacz_5	aaacggggatactgacgaaa	
Lacz_6	tgtgaaagaaagcctgactg	
Table 3: Oligos for ChIRP		

ChIRP-seq data analysis

Raw reads (100 bp paired-end) were quality trimmed (Trimgalore) and uniquely mapped to the mm9 reference genome using Bowtie2 (Langmead and Salzberg, 2012). Peak calling (ChIRPed DNA versus input) was performed with MACS 2.0.10 in paired-end mode with q-value cutoff < 0.05. One tag from each unique position was examined to eliminate peaks generated by clonal amplification. For peak calling in repetitive elements, reads with secondary alignments (allowing up to two alignments per properly-paired read) were also included.

Four independent LncHSC-2 (DT) ChIRP-seq experiments were performed: two consisted of non-overlapping "even" and "odd" probe sets (split-probe strategy) and two consisted of all combined probes. An additional negative control was generated by performing ChIRP-seq with probes targeting the bacterial *LacZ* gene. Input libraries were created for each sequencing run (n = 2).

The ChIRP-seq analysis workflow consisted of the following steps: (1) Find concordant regions: Identify all possible regions of overlap among significant peaks called in the four DT and control LacZ experiments. (2) Remove non-specific regions: Exclude regions covered by LacZ peaks and regions not covered by concordant peaks in the two split-probe experiments. Require remaining regions to be covered by peaks in at least one of the two complete probe set experiments. (3) Select consensus peaks: Merge any overlapping regions, select the first intersecting peak from one of the two complete probe set experiments. Exclude the consensus peak if it has any overlap with LacZ peaks.

Co-enrichment Analysis: Over-representation of LncHSC-2 consensus peaks (compared to LacZ peaks) in chromatin marks, TFBS, and regulatory elements was assessed by one-tailed Fisher's Exact Test with Benjamini-Hochberg multiple testing correction. Similarly, a two-tailed Fisher's test was performed to assess enrichment and depletion of LncHSC-2 peaks (compared to LacZ peaks) in repetitive sequences obtained from the outputs of RepeatMasker mouse mm9 assembly downloaded from UCSC website (http://hgdownload.soe.ucsc.edu/goldenPath/mm9/database/).

Mouse Phenotype Analysis: Predictions of cis-regulatory function of lincHSC-2 binding sites in Mouse Genome Informatics (MGI) Phenotype ontologies were performed with the GREAT 2.0 web portal utilizing the region-gene association rule: basal + extension: 10-kb upstream, 10-kb downstream, 20-kb maximum extension. Terms with a binomial test false discovery rate q-value < 0.05 are reported.

Motif Analysis: Sequences of LncHSC-2 (DT) consensus peaks within +/- 100bp around peak summits were extracted, repeat masked, and motifs analysis against these peaks was performed using DREME (Bailey, 2011). Only motifs with an E-value < 0.01 and p-value < 1.0e-006 were reported. Enriched motifs were searched against known motifs (JASPAR CORE vertebrates 2014 and mouse UniPROBE) using TOMTOM (Gupta et al., 2007). Matches with a q-value < 0.05 were considered significant.

Hierarchical Clustering: We selected the subset of 78 DT peaks co-occupied by one or more of the following factors in the given cell types: LT-HSC (UMR, H3K4me3 peaks), bone marrow (Enhancer regions, H3K4me3 and H3K27ac peaks, CTCF and PolII binding sites), HPC-7 hematopoietic stem/progenitor cells (Erg, Fli, Meis1, and Pu binding sites), E2a motifs (identified with DREME), CpG island sequences. After meeting initial selection criteria, the peaks were also assigned variables reflecting their association with Refseq transcripts (e.g., promoter, intragenic, or intergenic). Treating overlap of DT peaks with a given factor or sequence element as asymmetric binary variables, the pair-wise distance between variables was calculated by Gower's similarity coefficient and the clustering computed by Ward's method, both using the R cluster package.

Statistics

Student's t-tests and 1-way ANOVAs were used for statistical comparisons where appropriate. Significance is indicated on the figures with the following convention: $*P < 0.05$, $**P < 0.01$, $***P < 0.001$.

References for Supplemental Material

Bailey, T.L. (2011). DREME: motif discovery in transcription factor ChIP-seq data. Bioinformatics *27*, 1653-1659.

Challen, G.A., Boles, N.C., Chambers, S.M., and Goodell, M.A. (2010). Distinct hematopoietic stem cell subtypes are differentially regulated by TGF-beta1. Cell Stem Cell *6*, 265-278.

Chu, C., Qu, K., Zhong, F.L., Artandi, S.E., and Chang, H.Y. (2011). Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Mol Cell *44*, 667-678.

Dahl, J.A., and Collas, P. (2008). A rapid micro chromatin immunoprecipitation assay (microChIP). Nat Protoc *3*, 1032-1045.

Goodell, M.A., Brose, K., Paradis, G., Conner, A.S., and Mulligan, R.C. (1996). Isolation and functional properties of murine hematopoietic stem cells that are replicating in vivo. J Exp Med *183*, 1797-1806.

Grant, G.R., Farkas, M.H., Pizarro, A.D., Lahens, N.F., Schug, J., Brunk, B.P., Stoeckert, C.J., Hogenesch, J.B., and Pierce, E.A. (2011). Comparative analysis of RNA-Seq alignment algorithms and the RNA-Seq unified mapper (RUM). Bioinformatics *27*, 2518-2528.

Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., and Noble, W.S. (2007). Quantifying similarity between motifs. Genome biology *8*, R24.

Jeong, M., Sun, D., Luo, M., Huang, Y., Challen, G.A., Rodriguez, B., Zhang, X., Chavez, L., Wang, H., Hannah, R.*, et al.* (2014). Large conserved domains of low DNA methylation maintained by Dnmt3a. Nat Genet *46*, 17-23.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature methods *9*, 357-359.

Li, R., Yu, C., Li, Y., Lam, T.W., Yiu, S.M., Kristiansen, K., and Wang, J. (2009). SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics *25*, 1966-1967.

Mayle, A., Luo, M., Jeong, M., and Goodell, M.A. (2013). Flow cytometry analysis of murine hematopoietic stem cells. Cytometry A *83*, 27-37.

Wang, L., Park, H.J., Dasari, S., Wang, S., Kocher, J.P., and Li, W. (2013). CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Res *41*, e74.

Wilson, N.K., Foster, S.D., Wang, X., Knezevic, K., Schutte, J., Kaimakis, P., Chilarska, P.M., Kinston, S., Ouwehand, W.H., Dzierzak, E.*, et al.* (2010). Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. Cell Stem Cell *7*, 532-544.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome biology *9*, R137.